Philosophia Scientiæ
Travaux d'histoire et de philosophie des sciences

19-1 | 2015
Logic and Philosophy of Science in Nancy (II)

A New Role for Data in the Philosophy of Science

Molly Kao

OpenEdition Journals

Electronic version
URL: http://journals.openedition.org/philosophiascientiae/1052
DOI: 10.4000/philosophiascientiae.1052
ISSN: 1775-4283

Publisher
Éditions Kimé

Printed version
Date of publication: 1 March 2015
Number of pages: 9-20
ISBN: 978-2-84174-707-8
ISSN: 1281-2463

Electronic reference
Molly Kao, « A New Role for Data in the Philosophy of Science », *Philosophia Scientiæ* [Online], 19-1 | 2015, Online since 23 April 2015, connection on 03 November 2020. URL : http://journals.openedition.org/philosophiascientiae/1052 ; DOI : https://doi.org/10.4000/philosophiascientiae.1052

A New Role for Data in the Philosophy of Science

Molly Kao
The University of Western Ontario (Canada)

Résumé : Il existe un problème de circularité de la mesure : la construction des théories requière des données fiables, mais obtenir des données fiables requière des dispositifs de mesure dont la construction requière une théorie. Je soutiens qu'une possible solution à ce problème peut être trouvée en adaptant l'épistémologie empiriste de Anil Gupta au contexte de la science. On peut considérer les données, non comme un fondement pour la théorie, mais comme jouant un rôle fonctionnel, celui de rendre licite des révisions de la théorie antérieure. Les données autorisent les scientifiques à accepter des énoncés scientifiques sous la condition des théories d'arrière-plan qui sont les leurs. Une autorisation inconditionnelle est obtenue quand les différentes théories de départ convergent sur la même conception au cours de l'expérimentation. J'explique cette idée en utilisant deux exemples, l'un relevant de la thermométrie, l'autre mettant en jeu les expériences sur les courants neutres faibles.

Abstract: There exists a problem of the circularity in measurement: construction of theories requires reliable data, but obtaining reliable data requires reliable measurement devices whose construction requires a theory. I argue that adapting Anil Gupta's empiricist epistemology to a scientific context yields a possible solution. One can consider the role of data not as providing a foundation for a theory, but as acting functionally, licensing revisions of a previous theory. Data provide scientists with entitlement to their claims conditional on their background theory. Unconditional entitlement is obtained when different starting theories converge to the same view over the course of experimentation. I explain this idea using two examples, one in thermometry and one involving experiments on the weak neutral current.

Philosophia Scientiæ, 19(1), 2015, 9–20.

1 Introduction

The problem of theory-ladenness in the philosophy of science has many manifestations. For instance, in the post-logical positivist years, one prevalent strategy for discrediting the strict distinction between observational and theoretical terms was to point out that one's observational experiences are affected by the theory one brings to bear on the experience (cf. [Hanson 1958], [Kuhn 1970], [Feyerabend 1981]). The problem I will be addressing is not that the phenomenology of specific perceptual experiences can differ depending on the theoretical background an observer possesses. Instead, I will focus on the following problem, which is closer to one raised by Pierre Duhem: the construction of theories requires reliable data, but acquiring reliable data often requires some kind of theory to construct an accurate measuring device. Given this circularity, scientists must be able to provide justification for how they come to certain conclusions based on the experimental data they obtain. I will call this the problem of measurement which is, of course, distinct from the measurement problem of quantum mechanics.

In his book *Empiricism and Experience*, Anil Gupta introduces a novel empiricist epistemology that he believes is able to overcome many of the problems facing classical empiricism [Gupta 2006]. The primary purpose of his book is to argue that we can think of the rational contribution of experience to knowledge as falling within the logical category of a function rather than being propositonal in nature, and that this conception can provide a robust notion of entitlement and justification for our knowledge. In this paper, I explore the possibility of applying Gupta's account to the structure of scientific theorizing.

I will investigate this possibility by studying two examples, one involving thermometry, and one concerning experiments attempting to detect the weak neutral current. I will argue that the data obtained in scientific experiments can be conceived of as playing a functional role in scientific theorizing, analogous to the role Gupta assigns to experience in his everyday epistemology. These examples are meant to play a few roles. The first is to address the problem of measurement mentioned above, and to argue that Gupta's epistemology can provide us with a structure of reasoning that yields good justification for the judgments made in such cases. Secondly, although to defend aspects of Gupta's account is beyond the scope of this paper, I hope that outlining a plausible way in which to conceive of measurement in scientific practice using Gupta's epistemology will lend weight to the view he espouses.

I will begin by providing an overview of Gupta's empiricism in section 2. I will outline how his epistemology can be applied to understand the role of data in scientific theorizing in section 3. I will then clarify these claims in section 4 by examining two cases. Section 5 will present some possible objections and responses, and I will conclude in section 6.

2 Gupta's reformed empiricism

Gupta's empiricism is born of a desire to respect the idea that "experience is the principal epistemic authority and guide" [Gupta 2006, 4] as well as the idea that any particular subjective experience can be produced in multiple ways. His goal is to provide a brand of empiricist epistemology that incorporates these ideas, while not succumbing to some of the pitfalls of classical empiricism. In particular, Gupta is concerned to reject the idea that what is given in experience is propositional in nature.

Gupta's suggestion is to consider experience as falling into the logical category of a *function*. Thus, we must take seriously the idea that an experience alone cannot justify perceptual judgments: for judgments to be considered rational, they must be made over a background conception of oneself and one's place in the world. A subject's conception of herself and her conception of the world are interdependent. One is licensed to certain judgments only when one holds a particular view of the world, but one's view of the world is constantly being changed by the experiences one is having. It is key that the rationality of making certain claims is relative to one's conception of one's relation to the world.

For our purposes, we can consider a view to be a combination of judgments, propositions and concepts. An experience provides a mapping from a view to perceptual judgments. Schematically, if e is an experience, and v is a view, the logical contribution of experience is a function Γ_e. $\Gamma_e(v)$ refers to a class of schematized propositions that the subject is licensed to make, given her view and the experience she has. Gupta presents it thus:

$$\text{View } v \Rightarrow (\text{Experience } e \Rightarrow \text{Perceptual judgments } \Gamma_e(v)).$$

Thus, experience merely provides a rational link between views and judgments, and a reasonable view must accompany an experience to produce a reasonable judgment. Moreover, even after undergoing a particular experience, a subject may not be wholly entitled to the judgments thus produced. To see this, consider $\Gamma_e(v)$ a class of propositions, containing the judgment Q. It may be that the subject holds view v and undergoes experience e, where $Q \in \Gamma_e(v)$, and to update her view, the subject merely has to add Q to v. However, adding Q to v may make v inconsistent. In this case, she is not entitled to the judgment Q, and she must revise her view accordingly.

However, there seems to be a problem: if the entitlement to judgments that comes from experience is purely conditional, how do we ever acquire categorical entitlement to certain statements? Gupta's solution draws on the idea of a revision process. He uses experience to generate a revision process on different views and then finds the core of agreement between those views that results.

Consider an ideally rational agent. He will initially hold a view, v_0, and will undergo a succession of experiences $\mathscr{E} = \langle e_0, e_1, \ldots, e_n, \ldots \rangle$. At each stage n, the logical role of experience e_n is given by Γ_{e_n}, a mapping from the view the subject holds at that point to a new view. Thus, at each stage, a subject revises his view according to the nature of the experience. Sometimes, the revision consists merely of adding a judgment Q; at other times, revision may be a much more involved procedure in which contradictions are discovered and discarded. Thus, a sequence of experiences $\mathscr{E}$ acting on initial view v_0 will generate a sequence of views, $\mathcal{V} = \langle v_0, v_1, \ldots, v_n, \ldots \rangle$. Gupta calls $\mathcal{V}$ the *revision sequence* generated by $\mathscr{E}$ and v_0, where $\mathcal{V}_n$ refers to the view at stage n. A sequence of views $\mathcal{V}$ is *stable* iff there is a stage n after which all subsequent changes result in a view that is fundamentally equivalent to $\mathcal{V}_n$, where this simply means that the views all provide basically the same account of the world. Gupta then introduces the notion of *virtual identity* between different views. Views v and v' are virtually identical ($v \approx v'$) if they are the same aside from minor differences caused by differences in initial views [Gupta 2006, 93]. Two stable revision sequences $\mathcal{V}$ and $\mathcal{V}'$ *converge* iff there is a stage n such that for all $m \geq n$, $\mathcal{V}_m$ is virtually identical to $\mathcal{V}'_m$.

Consider again an ideally rational being. She would be able to conceive of all the views that would be admissible as starting points. She would also be able to imagine the effects of a sequence of experiences $\mathscr{E}$ on each of those views. Take $\Pi_{\mathscr{E}}$ to be the function taking admissible views to revision sequences and call it a revision process. Then the revision process $\Pi_{\mathscr{E}}$ is strongly convergent iff there is a stage m where all the revision sequences become virtually identical. "*Strongly convergent processes generate absolute rational obligations*" [Gupta 2006, 98]. If all admissible views were to converge to virtually identical ones, a rational agent would be obligated to hold that view. Furthermore, even if the sequences of views generated by a revision process disagree on details, whatever core of agreement exists would also impose rational obligations on the agent. Thus, convergence is what provides categorical entitlement in Gupta's account. In the next section, I will reinterpret these ideas in the context of epistemology of science, and in particular, apply it to the problem of measurement.

3 A new role for data

In the following, I will refer to "data" in a way that closely follows the definition given by Bogen & Woodward in [Bogen & Woodward 1988].[1] However, for my purposes, it is not important to take data as a contrast to phenomena, and I will speak of data simply as providing evidence for changing our theories. I take "data" to refer to the individual results of observations or measurements. Woodward describes data roughly as what registers on a measurement or recording device in a form which is accessible to the human perceptual system, and to public inspection [Woodward 1989, 394]. Data need not be strictly

1. This distinction is discussed further in [Woodward 1989, 2000] and [Bogen & Woodward 1992, 2005].

numerical, for they can refer to the results of a number of different detection procedures such as temperature readings, scores on psychological tests, and spark detector photographs [Bogen & Woodward 2005].

We now return to the problem of measurement and how it may be resolved. My suggestion is to adapt Gupta's epistemological structure to apply to scientific investigation. The adaptation is a natural one: we simply take a scientist's background theory to play the role of Gupta's "view", and consider the rational role of data as providing entitlement to certain claims, conditional on the scientist's background theory and the data. Data on their own do not license us to make judgments, just as experience itself does not license judgments. Rather, one must have a background theory, and data play the role of providing a mapping from the original background theory to a revised theory. A series of experiments is analogous to a sequence of experiences, and the course of experimentation induces a succession of theories, each time revised according to the data obtained, just as a sequence of views is revised according to an agent's experiences. In ideal cases, convergence between different initial views ensues. In such cases, we can explain why data are not cripplingly theory-laden. Although data do not provide justification for claims apart from a theory, this is not to say that there cannot be justification for any claims. Rather, unconditional justification arises when there is convergence between different theories.

One aspect of Gupta's epistemology that makes it appropriate for an epistemology of science is his claim that what constitutes perceptual beliefs can shift, depending on the context. Perceptual judgments refer to those judgments that are immediately yielded by an experience in conjunction with a view, and need no further justification. However, in cases where one's view is challenged, that immediate entitlement may no longer exist, and so the judgment would not count as perceptual. For instance, the claim "The apple is red", may count as a perceptual judgment in normal circumstances, but if one is then told that she is in a room with unusual red lighting, the entitlement to that claim is undermined. This is in line with Prajit Basu's claim that in science, what counts as "raw data" depends on both the context of the experiment and the background knowledge being taken for granted by the scientific community [Basu 2003]. He argues that evidence for a theory is constructed within that theory from (raw) data, but the move from data to evidence can occur at many levels. This can easily be explained in Gupta's framework. Certain data license one to certain judgments only in conjunction with a background theory, and in the absence of shared views in scientists' background theories, it will be impossible for the same data to have the same rational import. In the next section, I will present some examples to make explicit the application of this framework.

4 Two examples

4.1 Measuring temperature

My first example will demonstrate the problem of testing measurement devices without circularly justifying their use. The history of thermometry is a long one, so I will focus on one stage, when reasonably reliable thermometers had been developed and it was standard to calibrate them according to the freezing and boiling points of water. It seems, though, that in order to determine the thermometer that most accurately gives the temperature between these points requires knowledge of the temperature of the substance being measured, i.e., an accurate thermometer. This example demonstrates how convergence between different views can yield categorical entitlement on certain questions. All my historical information will be drawn from the excellent and detailed study presented by Hasok Chang in his *Inventing Temperature* [Chang 2004].

One important debate in thermometry was focused on the best choice of fluid for determining the temperature of a substance, where two main contenders were mercury and air. Jean-André De Luc tried to solve the problem by using the "method of mixtures", in which different proportions of water at the boiling and freezing points were mixed together. Thus, a mixture of 75% boiling to 25% freezing water was assumed to have a temperature of 75°C. As a proponent of the mercury thermometer, De Luc wished to show that it would yield measurement results that were very close to the calculated value, and his experimental results were indeed quite close. De Luc also conducted comparative experiments, where he compared the calculated degree of real heat with thermometers filled with substances other than mercury. The mercury thermometer yielded the most accurate results. De Luc thus concluded that the mercury thermometer gave the best approximation to the real temperature.

I claim that it is possible to understand the role of the data in De Luc's theorizing as entitling him to certain judgments, which he then had to incorporate into his view. Presumably, each measurement result allowed De Luc to infer something specific. For instance, De Luc may have obtained the first measurement result for water calculated to be at 75°, and concluded that "The mercury thermometer shows 74.7° when the water is at 75°." A reading of 37.0° from the thyme oil thermometer for water that was calculated to be 40.0° allowed him to conclude "The thyme oil thermometer reads 37.0° when the water is at 40.0°." A reading for the alcohol thermometer allowed him to conclude "The alcohol thermometer reads 33.7° when the water is at 40.0°." Incorporating these into his view would then allow him to conclude that "The thyme oil thermometer is more accurate than the alcohol thermometer."

In providing the structure of De Luc's theorizing, it is important to realise what was part of his view: one important feature was that he believed that the method of mixtures was an accurate way of preparing water at a certain temperature. Given this, and given the data that he obtained, he was justified in accepting the judgments that he did. As it turns out, this assumption (and several others) were later challenged; when this occurred, De Luc's entitlement to the judgments about the readings of the mercury thermometer was undermined, and further work had to be done.

I will now consider some of the subsequent developments in thermometry and explain how these fit naturally into the framework being put forth. It will be necessary to start by outlining the admissible views, in this case, opinions about what type of thermometer provided the most accurate measure of temperature. These opinions were often substantiated by specific theories of heat and molecular motion. The first view under consideration was De Luc's. A different view that issued direct challenges to some of the aspects of De Luc's view was the newly emerging caloric theory of heat. This theory held that caloric was a subtle fluid that was either the cause of heat, or heat itself. On one variety of this theory, the amount of caloric in a substance is a product of the substance's capacity for caloric and its temperature. Thus, if a certain body preserved its heat content (the amount of caloric) but its capacity for caloric was raised, its temperature would go down. One of the theory's most prominent proponents was John Dalton, who was able to challenge De Luc's use of the method of mixtures. Dalton argued that when hot and cold water were mixed, there was a slight decrease in overall volume. According to Dalton's view, this meant that there was less room for caloric to fit into that water, and so there would be a decrease in heat capacity, or capacity for caloric. Thus, the temperature would go up, with the result that De Luc's "method of mixtures" calculations would give a temperature that was actually too low. This is analogous to the phenomenon Gupta describes as the shift in what constitutes perceptual judgments, depending on the context. When De Luc was experimenting, he obtained certain thermometer readings, and was licensed straightforwardly to the judgment "The mercury thermometer shows 74.7° when the water is at 75°." By challenging De Luc's view, Dalton also challenged his entitlement to that judgment.

A view that emerged later with respect to the question of the most reliable thermometric fluid was that of Henri Victor Regnault. Rather than designing his experiments based on a particular theory of heat, he carried out a series of experiments intended to test the concept of "comparability". As Chang explains:

> If a thermometer is to give us the true temperatures, it must at least always give us the same reading under the same circumstance. [Chang 2004, 77]

Thus, Regnault's view regarding the best thermometer did not include a theory of heat or molecular motion; instead, it included a principle about what constituted "best" when judging the accuracy of thermometers. Regnault also did not begin with a view that took either mercury or air thermometers to be superior; his initial view was neutral on the question, and he revised it in light of the data he obtained.

Regnault proceeded to carry out a number of experiments to test comparability. One of his tests was designed to compare the values given by mercury thermometers for the same substance: a variety of mercury thermometer readings for the same substance were compared to a reading from an air thermometer. As it turned out, different mercury thermometers made with different types of glass gave different readings. Here, different data allowed Regnault to form judgments such as "The mercury thermometer made with ordinary glass reads 149.80° when the air thermometer reads 150°", and "The mercury thermometer made with crystal reads 150.40° when the air thermometer reads 150°."[2] When the judgments from the data were incorporated, he was then able to draw a conclusion about the comparability of those thermometers, or the lack thereof. Experiments on different types of air thermometers yielded much more positive results. Such data, along with Regnault's commitment to the idea that comparability provided the best standard for evaluating thermometers licensed him to certain judgments. Anyone who shared the idea that comparability was an important standard, but held that mercury thermometers were the superior instruments, would have to revise their view in light of such data.

I argue that we can see this progression as an instance of convergence of views, providing us with categorical entitlement to the idea that the air thermometer was the most accurate of the choices, given the data that were produced. We began with different views as to what constituted the best thermometric fluid. A series of data caused a change in all of those views. The core of agreement at the end of the series of experiments was that the air thermometer provided a more reliable measure of temperature than other types of thermometers. If a different sequence of data had been obtained, working scientists would have been rationally obligated to a different set of beliefs that constituted the core. For instance, if Regnault's tests on the mercury thermometer had shown that they satisfied the standard of comparability, he would have had to revise his view that was neutral between air and mercury thermometers. This would have been in agreement with De Luc's views, providing a different example of convergence between two different initial views, revised in the light of experimental data.

It is important to note that what did not result from this data was convergence with respect to the underlying theoretical views. The experiments did not license scientists to make judgments that could affect those views. However, even without a definitive theoretical account, the convergence of several views regarding thermometric fluids entitled scientists to a core of agreement. Other experiments and other data would have to be used in order to come to a consensus on different questions regarding heat. This shows that the idea of a "view" in science can be as fine-grained as we wish to take it; scientists may have a view with respect to a particular question, e.g., the best fluid for thermometry, or they may have a view on a more theoretical ques-

2. Data is taken from [Chang 2004, 80, Table 2.4].

tion, e.g., what constitutes heat. There will certainly be overlap between such issues, but I take it as a virtue that we can conceive of views so flexibly, since scientists are often concerned with particular questions, and it is desirable to have an account of scientific reasoning that can apply in all those contexts. When convergence is achieved on a particular question, then scientists are entitled to that view, although there may have been erroneous elements influencing the data acquisition. Convergence here shows that although certain theories may have been involved in constructing views, and perhaps in constructing measurement devices, agreements can ensue, eliminating the vicious circularity. If we conceive of data as functional rather than a foundation on which to build further theory, its role becomes unproblematic.

4.2 Detecting weak neutral currents

In the last section, I discussed an example where the data licensed experimenters to judgments whose contents straightforwardly stated numerical values obtained in the measurements. In this section, I will discuss the use of photographs of bubble chambers in order to detect weak neutral currents.[3] I will not treat it in great detail, but it will be useful to see how this account of the role of data can easily accommodate data that is non-numerical in character, and is thus not easily thought of as yielding a judgment that is merely its sentential expression.

In 1973, physicists were able to detect weak neutral currents through a series of experiments. At CERN, this was done by firing a neutrino beam into a bubble chamber and taking photographs of the results. According to Bogen & Woodward,

> The data obtained at CERN consisted of approximately 290,000 bubble chamber photographs of which roughly 100 were thought to provide evidence for the presence of neutral currents. [Bogen & Woodward 1988, 315]

In this case then, the data consisted of a set of photographs. In order to understand the role of this data in theorizing from our Guptan perspective, we consider these photographs as licensing certain judgments, given a particular background theory. I will argue that this is precisely the role that the photographs play.

One of the factors that made this experiment problematic was the fact that there are other interactions that mimic the behaviour of neutral currents. When a neutrino enters the chamber, either a charged current interaction or a neutral current interaction may occur. Both types of interactions produce a shower of tracks, but a charged interaction will also produce a high energy

3. My discussion here will be simplified, but will follow the example as laid out in [Bogen & Woodward 1988, 2005; Woodward 1989].

muon, which leaves its own track. However, the neutrino also interacts with the chamber in such a way that it may produce neutrons which interact with other particles in the bubble chamber, imitating a neutral current interaction. This phenomenon thus produces false positives for neutral current interactions. It was necessary to have accounted for this "neutron background" in order to assert that the photographs were genuinely evidence for neutral currents. There were experimenters who used different techniques to establish an upper bound on this neutron background. For those holding the view that this upper bound was correct, the collection of photographs allowed them to judge that some photographs were displaying weak neutral current interactions. Without the belief that such an upper bound existed, scientists would not be licensed to this claim.

I submit that here, unlike in the temperature measurement examples, there is no clear way to conceive of either a particular photograph or the collection of photographs as yielding a statement that obviously captures their content. The judgments scientists were licensed to make from individual photographs were something like "This is a picture of blobs and flashes", which could then be incorporated into their view that would allow them to infer that the pictures represented a possible neutral current interaction. However, such a statement is not the same thing as the image itself, and so it is more reasonable to think of the photographs (the data) as allowing certain judgments to be made. The collection of photographs would then license a judgment such as "There are 100 photos displaying these patterns of blobs and flashes." This, in combination with the view that held that the upper bound of the neutron background was lower than what was found in experiments licensed inferences about neutral currents. Thus, we can think of the photographs as allowing for an inference, given someone's view of the apparatus and other theoretical factors.

5 Some possible objections

One obvious objection to this account of scientific theorizing is that it is not realistic to expect different scientific theories always to yield convergence. It may happen that a long string of experiments never causes multiple views to fully converge. Moreover, the opposite scenario can and has occurred. Scientists have believed that they possess categorical entitlement to a view, or at least to a set of beliefs, only to discover at some later date that these views have been incorrect. I grant both points, but I do not think that these are truly objections to the framework that has been set out. In the first case, it is perfectly natural that multiple views will not necessarily converge—the lack of convergence is very often what incites scientists to design more experiments and collect more data. However, even in cases where views do not become virtually identical, they will often contain a core of agreement. In these cases, scientists are categorically entitled to those statements, even though both general views remain rational. As for the second case, this also makes good sense given the history of science. Because we are not ideal beings, scientists cannot possess all possible theories at the beginning of the course of experimentation. The "correct" view may not even be in the realm of possibility given the theoretical resources at hand. Thus, there are cases where scientists' available views have converged, but new data (combined with their current theories) forces them to make judgments which render their previous views inconsistent, and therefore obliges them to revise their views. This process may generate several new views, and the search for convergence continues. This explains why scientists are rational to continue their scientific inquiry in this way.

Another problem one might raise is that this account does not necessarily make the revision rule algorithmic, either for the commonsense view, or for scientific enterprises. However, I would argue that the rule of revision need not be perfectly algorithmic, nor must it be unique. We can grant that there are different ways to make rational decisions without denying that the epistemological role of data in scientific investigations are to allow scientists to make judgments. Certainly more work would have to be done in order to outline how we know what judgments we are licensed to make, given a view and data. I suspect that these factors would be highly contextual, and depend on the phenomena under investigation.

A different concern might be that even if Gupta's account of empiricism works for the commonsense view, it does not extend so easily to science since the nature of data differs phenomenologically from the nature of experience. Experience just is a phenomenological occurrence, whereas data goes far beyond a subject's phenomenology. The relation between experience and allowable perceptual judgments seems much more direct than the relation between data and the judgments one is entitled to due to that data. While it is true that data is quite different from "experience" phenomenologically speaking, this does not negate the possibility of their playing the same logical role. On Gupta's account, the "direct awareness" aspect of experience is merely a byproduct of our constitution. The *character* of experience is not the primary contributor to the force of rational experience. Thus, it is not problematic that data do not share this subjective character. What matters is that they play similar roles in their respective realms of reasoning.

6 Conclusion

I have argued in this paper that Gupta's novel empiricism can plausibly be used in the philosophy of science to account for the problem of circularity in measurement. In particular, I hope to have shown that conceiving of data as playing a functional role, and providing hypothetical entitlement to certain judgments, makes sense of the epistemological structure of scientific investigations. By assigning data this role, scientists can gain categorical entitlement to judgments when convergence occurs over possible views, or even over small parts of views. The examples I have given are merely a small start—it would require far more work to show that the same epistemological picture could be applied in more complex situations. Nevertheless, it appears to be a promising account of what justifies scientists in some of their claims, and I believe could be fruitfully extended.

Bibliography

BASU, Prajit K. [2003], Theory-ladenness of evidence: A case study from history of chemistry, *Studies in History and Philosophy of Science – Part A*, 34(2), 351–368, doi:10.1016/S0039-3681(03)00022-0.

BOGEN, James & WOODWARD, James [1988], Saving the phenomena, *The Philosophical Review*, XCVII(3), 303–352.

—— [2005], Evading the IRS, in: *Idealization XII: Correcting the Model: Idealization and Abstraction in the Sciences*, edited by R. Jones, M. & N. Cartwright, The Netherlands: Rodopi, 233–268.

BOGEN, Jim & WOODWARD, Jim [1992], Observations, theories and the evolution of the human spirit, *Philosophy of Science*, 59(4), 590–611.

CHANG, Hasok [2004], *Inventing Temperature: Measurement and Scientific Progress*, New York: Oxford University Press.

FEYERABEND, Paul K. [1981], *Realism, Rationalism, and Scientific Method: Philosophical Papers*, vol. 1, New York: Cambridge University Press.

GUPTA, Anil [2006], *Empiricism and Experience*, New York: Oxford University Press.

HANSON, Norwood Russell [1958], *Patterns of Discovery: An inquiry into the conceptual foundations of science*, New York: Cambridge University Press.

KUHN, Thomas S. [1970], *The Structure of Scientific Revolutions*, Chicago: University of Chicago Press, 2nd edn.

WOODWARD, Jim [1989], Data and phenomena, *Synthese*, 79(3), 393–472, doi: 10.1007/BF00869282.

—— [2000], Data, phenomena, and reliability, *Philosophy of Science*, 67 (*Proceedings*), S163–S179.